A Real-Time Online Decision Support System for Intermodal Passenger Travel

By
Mengqi Hu, PI
Department of Industrial and Systems Engineering
Mississippi State University
PO Box 9542, Mississippi State, MS 39762

Yao-Jan Wu, Co-PI
Department of Civil Engineering and Engineering Mechanics
University of Arizona
1209 E 2nd St., Tucson, AZ 85721

Yang Chen, PhD Student
Department of Industrial and Systems Engineering
Mississippi State University
PO Box 9542, Mississippi State, MS 39762

Shu Yang, PhD Student
Department of Civil Engineering and Engineering Mechanics
University of Arizona
1209 E 2nd St., Tucson, AZ 85721

NCITEC Project No. 2013-13

Conducted for

NCITEC

September 2015

DISCLAIMER

ABSTRACT

The transportation system in the United States is disjointed and inefficient as a result of the different transportation modes in use and their respective industries which have developed independently. In addition, public transportation is not well used in passenger trips compared to other developed countries. For example, public transportation accounts for 20% of passenger trips in large U.S. cities compared to 50% in Europe. Also, development of the passenger intermodal transportation system has lagged behind development of the freight transportation system.

To improve utilization of intermodal transit and efficiency in the U.S, we developed an intelligent decision support system for passenger travel decisions using real-time general transit feed specifications (GTFS) data. In our system, an automatic data collection strategy was created to collect GTFS and flight data across different platforms, and an "all-in-one" database was designed to store the data. The database was used to: 1) construct intermodal transit networks using a "node-link" scheme, and 2) estimate travel time and travel time reliability for links and transit routes. Using this real-time data, a data-driven travel decision model was developed to determine the best route based on passenger preferences. Several chance constraints were added in the decision model to guarantee the reliability of the travel route under uncertainties. Additionally, a user-friendly interface was developed in Python to allow travelers to plan their trips, and a geographic information system (GIS), Google Earth, was employed to allow users to visualize the optimized route options.

The proposed system was validated using real-time GTFS data collected in Tucson, AZ, and Boston, MA. This validation demonstrated that the system can determine optimal travel routes for passengers. In addition, three sets of sensitivity analysis experiments were developed to investigate three model considerations: 1) the effect of chance constraints on path choice, 2) the effect of confidence levels on path choice, and 3) the difference between weekend and weekday travel planning. The results suggested that the optimal anticipated travel time increases with an increasing on-time arrival confidence level, and walking is preferred by passengers instead of transferring buses during peak hours. As an example, approximately 30% additional time serves as a reference for allocating travel buffer time to ensure a higher on-time arrival confidence level for transit trips to the Tucson International Airport.

ACKNOWLEDGEMENTS

This research is supported by fund from Department of Transportation via National Center for Intermodal Transportation for Economic Competitiveness under contract number DTRT12GUTC14. The authors would like to thank Sun Tran and Massachusetts Bay Transportation Authority (MBTA) for data support. Special thanks to Michael Beel for real-time transit data collection. We are grateful to Metro Transit – Saint Louis for student training. We really appreciated Bernadette M. Marion’s efforts to support our students’ training. We would like to thank the following students for data collection/analysis and English writing assistance: Ahmad Fakhouri, Amin Ariannezhad, Anthony Huynh Giang, Ryan Brown, Joel Gerardo Amarillas and Payton Cooke.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUTION

Travelers across the country have been encouraged to switch travel modes from private vehicles to public transportation. A decrease in vehicle traffic can lead to carbon emission reductions and alleviation of traffic congestion. Transit service improvements including accessible stops, reliable trips, and mobility provisions provide travelers with incentive to switch to transit [1]. In addition to the development and maintenance of transit facilities, emerging transportation management information technologies have also been developed. Many transit agencies in the U.S. and Canada have adopted these technologies and created advanced public transportation systems (APTS) to improve their operational efficiency and service safety [2,3]. Therefore, transit systems are increasingly playing an important role in surface transportation.

One of the primary goals of transit operation is to increase ridership. Transit agencies have utilized APTS to release up-to-date transit information (e.g. changes in routes or timetables). A recent survey [4] suggested that the overall satisfaction and usability of transit systems would be improved by providing passengers with up-to-date transit information. Moreover, this information could also help transit passengers plan their trips to save travel time and expenses. Related trip planning applications can be found in [5–7]. The benefit of releasing up-to-date transit information is dependent on the quality of data collection. Conventional transit data collection approaches primarily include manual collection at pre-defined points (e.g. bus transfers and terminals) and transit-related surveys. Additionally, various technologies have been applied to automatically collect transit data. One of these technologies is GPS, which assists transit agencies and passengers in tracking and locating transit fleets in real-time. The ease of transit data collection using GPS ensures that transit information is up-to-date, simplifies the trip planning process, and makes planned trips reliable.

Reliable trip planning has been studied with the objective of saving transit passengers travel time and cost. Most transit agencies have developed their own trip planning systems for their riders. For example, Sun Tran [8] has developed a trip planning system based on its scheduled trips in Tucson, Arizona. Another trip planning framework is OneBusAway, [9] which integrates transit networks and schedule information in five cities. Although these trip planning systems can be easily accessed using webpages or apps on mobile devices, few academic studies have addressed their trip planning algorithms and models. Yan et al. [10] sought to minimize the expected values of random schedule deviations (specifically the sum of expected values), and their results showed that optimal scheduled travel time depends on bus drivers' schedule recovery behavior and decision makers' scheduling philosophies. In order to minimize wait times at transfer stations, Shafahi and Khani [11] formulated the issue of transit trip planning as a mixed integer programming model and presented a genetic algorithm approach to more efficiently find optimal transit trips in larger networks.

Beyond transit trip planning, several previous studies on general transportation trip planning are also summarized below. Travel time is one of the most important criteria when planning a trip. Travelers are commonly concerned with how long will their trip take. Thus, travel time is an essential factor for finding optimal trip paths. Recent studies [12–14] also suggested that travel time reliability (also known as travel time uncertainties) could be considered as "risks" of late arrival, because late arrivals caused by travel time variations would excessively inconvenience travelers (e.g., missed flights). Thus, travelers would desire to know how reliable their planned trip path would be. For example, airline passengers appear to depart early to create a "safe margin" to minimize the chance of missing their flight. This safe margin is usually called buffer time. Appropriate buffer time for different traveler routes could be provided to support better trip decisions depending on the departure time period on a specific day. Therefore, both travel time and travel time reliability are essential factors for trip planning.

2. OBJECTIVE

The main research goals of this project are to improve the efficiency of intermodal passenger transportation, improve the utilization of public transportation modes, and reduce transportation cost and travel time for passengers.

3. SCOPE

To attain the objectives listed above, an efficient and effective database was designed to provide an “all-in-one” data platform for storing static and dynamic information of different transportation modes (e.g., bus, flight, etc.). An automatic data collection approach was developed to collect real-time data and store it in the database.

Next, to help passengers reliably arrive at the airport on-time, both bus transit measures (travel time and travel time reliability) were incorporated in the proposed chance constrained decision model to obtain optimal paths at a predefined on-time arrival confidence level.

Finally, based on the database and the travel decision model, a user-friendly interface was developed to allow passengers to search for and schedule their travel plans. To visualize the optimized route options, Google Earth was used to display different routes.

4. METHODOLOGY

4.1 Data Collection

4.1.1 GTFS Data

The transit data followed the general transit feed specification (GTFS) format and was reported to Google in two types: GTFS static and GTFS real-time, with definitions and applications of each shown below.

- "The GTFS defines a common format for public transportation schedules and associated geographic information" [15]. GTFS has several fields, including "stop time", which refers to transit scheduled timetable, "route", "trip", and "stops", which contain the existing transit facility information. These fields are necessary to calculate the earliness/lateness of buses at stops. The field "shapes" provides the spatial references to locate transit fleet on roadways. Time-space diagrams can be plotted using the travelled mileages transformed from the real-time transit fleet locations.
- "GTFS real-time is a feed specification that allows public transportation agencies to provide real-time updates about their fleet to application developers. It is an extension to GTFS, an open data format for public transportation schedules and associated geographic information" [16]. It provides not only transit fleet information but also estimated trip status. Because this study aimed to measure reliability at the stop-level, only the real-time transit fleet locations contained in the field "VehiclePosition" were used. The update frequency of the GTFS real-time data was 30 seconds.

Figure 1 depicts the GTFS data flow where GTFS static and GTFS real-time can be accessed by different approaches. Transit agencies usually implement and release GTFS static data files on static webpages. Anyone can directly download these files from the webpages. GTFS real-time data is updated at a specific time interval and released through web service definition language (WSDL) technology. Users usually have to develop computer programs to grab the real-time data. Since GTFS real-time is encrypted with "protocol buffer" [17], GTFS real-time data should be decrypted before further processing. After downloading and grabbing the GTFS static and GTFS real-time data files, a customized computer program is developed to parse these files and import the parsed data into a database. Figure 2 lists the attributes in GTFS real-time data.

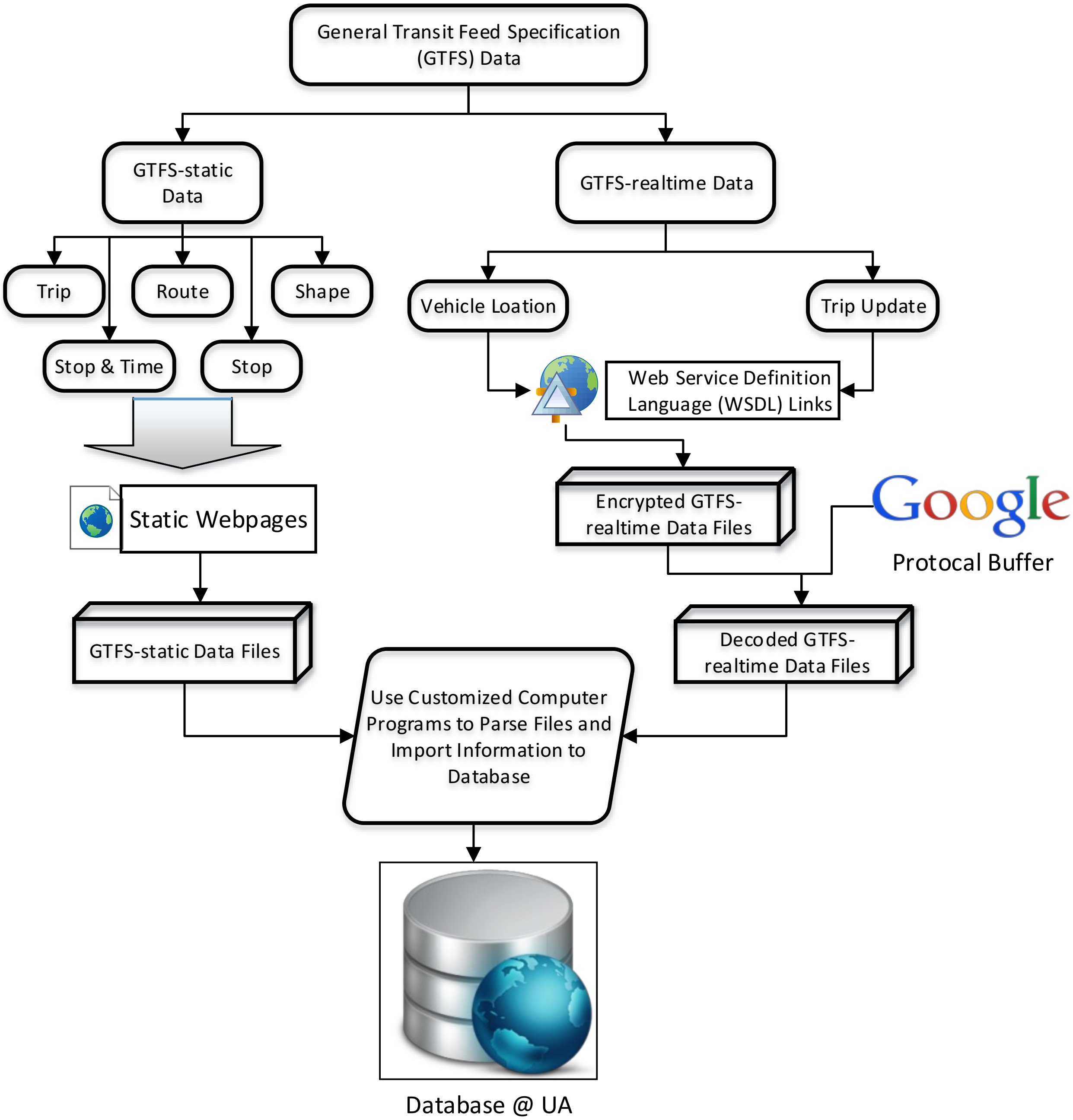


FIGURE 1 GTFS data overview

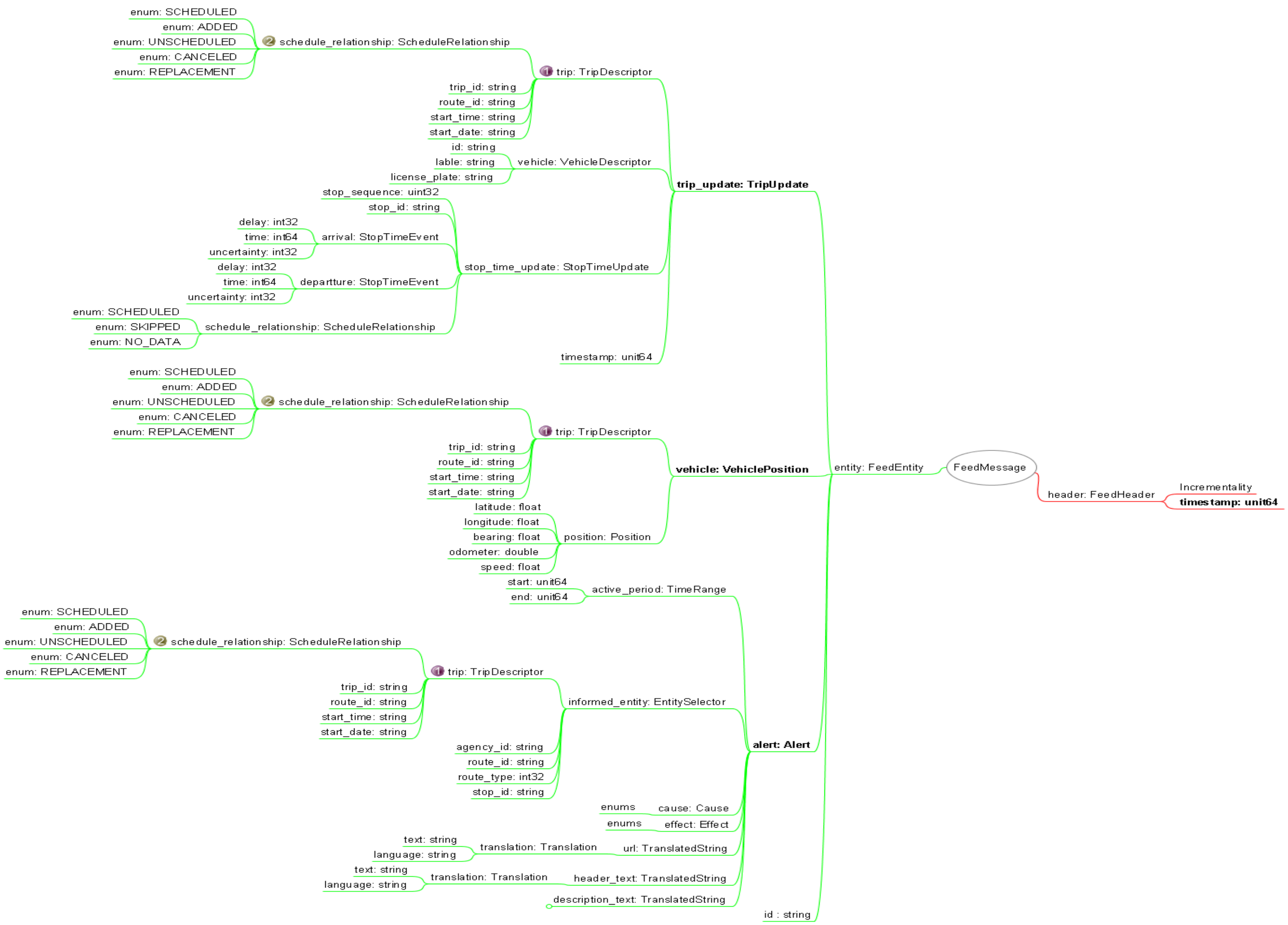


FIGURE 2 Real-time data attributes

4.1.2 Transit Agency Data

Transit service data assists transit agencies in making decisions with respect to transit operations and planning. Manually collecting transit service data has been a popular approach in the past several decades. Recently, emerging techniques have allowed decision makers and researchers to automatically collect transit service information. For instance, GPS is able to locate transit fleets in real-time. The automated vehicle location (AVL) system is built based on GPS. The transit service data collected from the AVL system contains not only fleet location information but also travel-related information (e.g. trip, route, and bus stop arrival time). However, specific transit agencies may define AVL data formats to satisfy their operation and planning requirements, without adhering to a common format. Google encourages transit agencies to follow the data format defined in GTFS to exchange and share transit service information. GTFS data is becoming more popular throughout the nation.

Figure 3 shows an overview of bus stops in Tucson, Arizona. This data was used to build a network for trip planning and path choice. The real-time transit fleet information was encrypted in the GTFS real-time format. Sun Tran, which manages transit service (including over 30 routes and over 2,000 bus stops) in Tucson, AZ, has implemented the two GTFS data formats and made them accessible to the public. Both types of GTFS data were collected from August 2014 to June 2015 and used in our study.

Two commonly used transit service measures were estimated using the GTFS data, including the mean value of link travel time (also known as stop-to-stop travel time) and transit service reliability. Transit service reliability is defined as the variance of link travel time. Travel time reliability is typically measured by time of day (TOD) and day of week (DOW) [18]. Sun Tran adopted two timetables (one for weekdays and one for weekends) because of the significant differences in transit demand between weekdays and weekends. Therefore, the transit service reliability is measured by TOD and weekdays or weekends. A dummy variable w is used to indicate either weekdays or weekends.

$$\overline{TT_{rn,l}^{tod,w}} = \frac{1}{K} \sum_{k=1}^{K} TT_{rn,l,k}^{tod,w} \tag{1}$$

$$\overline{TTR_{rn,l}^{tod,w}} = \frac{1}{K} \sum_{k=1}^{K} \left(TT_{rn,l,k}^{tod,w} - \overline{TT_{rn,l}^{tod,w}} \right)^2 \tag{2}$$

where TT and TTR are average and variance of link travel time given rn, l, tod, and w;

rn represents the route number;

l represents the lth link on Route rn;

w is 0 and 1 for weekends and weekdays, respectively.

k is the k^{th} estimated link travel time given rn, l, tod, and w;

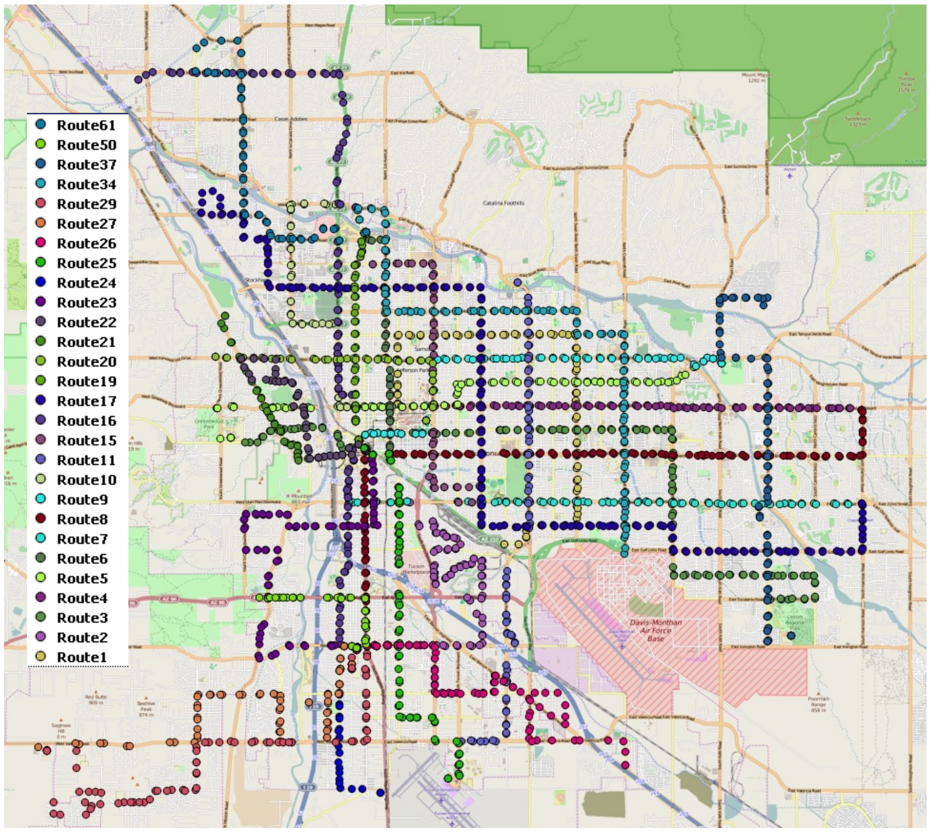


FIGURE 3 Bus stops and transit network in Tucson

(Background image is from the OpenStreetMap)

4.2Database Schema

According to the attributes of GTFS static and GTFS real-time, Figure 4 shows the database schema design. GTFS static data includes trip, route, stop&time, stop, and shape information. The primary keys are labeled in tables. All tables are connected by primary keys. For example, a trip is connected with the primary keys route ID and shape ID in Route and Shape tables, respectively. Figures 5-6 give examples of tables in database.

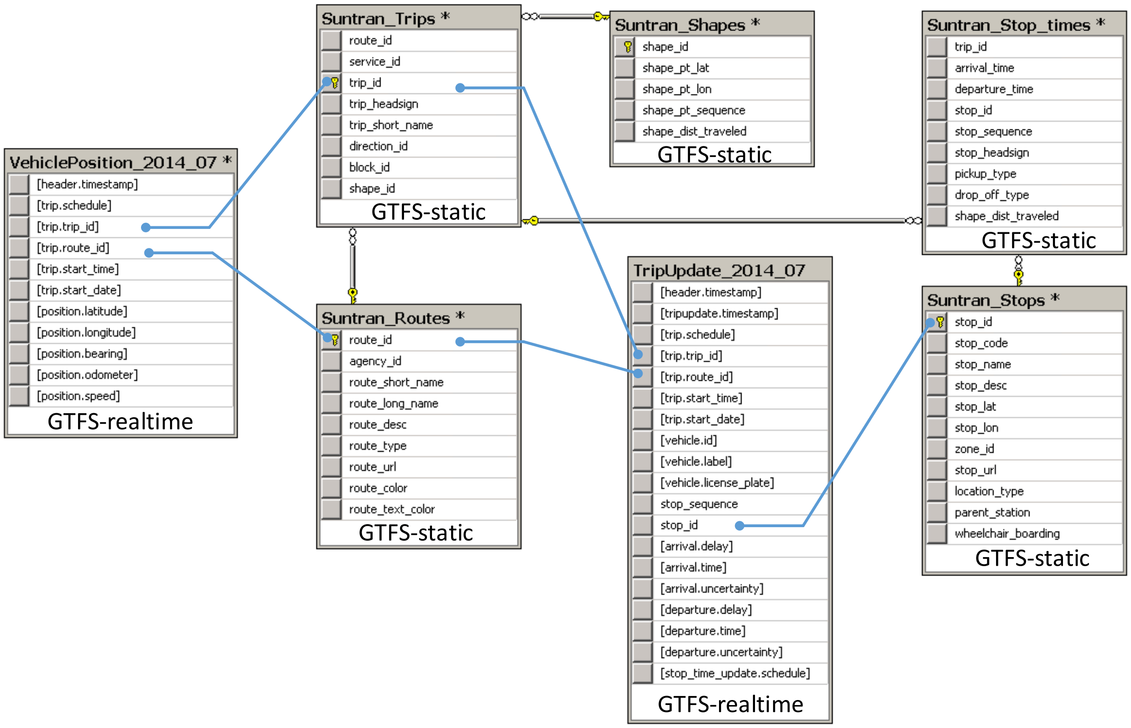


FIGURE 4 Database schema design

	route_id	agency_id	route_short_name	route_long_name	route_desc	route_type	route_url	route_color	route_text_color
1	11670	SunTran	1	Glenn/Swan	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_1.pdf	FFD457	000000
2	11617	SunTran	10	Flowing Wells	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_10.pdf	00703C	FFFFFF
3	11618	SunTran	101X	Golf Links-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_101X.pdf	003B7B	FFFFFF
4	11619	SunTran	102X	Northwest-UA Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_102X.pdf	C5913A	FFFFFF
5	11620	SunTran	103X	Northwest-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_103X.pdf	957CA0	FFFFFF
6	11621	SunTran	104X	Marana-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_104X.pdf	451F03	FFFFFF
7	11622	SunTran	105X	Foothills-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_105X.pdf	E00085	FFFFFF
8	11623	SunTran	107X	Oro Valley-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_107X.pdf	5B57A6	FFFFFF
9	11624	SunTran	108X	Broadway-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_108X.pdf	DA9B3C	FFFFFF
10	11625	SunTran	109X	Catalina Hwy-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_109X.pdf	4D4A8D	FFFFFF
11	11626	SunTran	11	Alvernon	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_11.pdf	7C5C36	FFFFFF
12	11627	SunTran	110X	Rita Ranch-Downtown Express	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_110X.pdf	00B1EB	FFFFFF
13	11628	SunTran	15	Campbell	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_15.pdf	00569C	FFFFFF
14	11629	SunTran	16	12th Ave./Oracle	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_16.pdf	00A3B4	FFFFFF
15	11671	SunTran	17	Country Club/29th St.	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_17.pdf	52BCC6	FFFFFF
16	11631	SunTran	19	Stone	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_19.pdf	00C200	FFFFFF
17	11632	SunTran	2	Pueblo Gardens	NULL	3	http://suntran.com/pdf/routes/AUG_14_rt_2.pdf	C4A4A5	FFFFFF

(a) Route table

	shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
1	35916	32.22265	-110.968297	1	0
2	35916	32.222707	-110.968297	2	0.006
3	35916	32.222822	-110.968297	3	0.019
4	35916	32.222853	-110.968297	4	0.023
5	35916	32.222968	-110.968297	5	0.035
6	35916	32.223078	-110.968297	6	0.048
7	35916	32.223194	-110.968298	7	0.061
8	35916	32.223229	-110.968298	8	0.064
9	35916	32.223345	-110.968298	9	0.077
10	35916	32.223488	-110.968298	10	0.093
11	35916	32.223544	-110.96803	11	0.119

(b) Shape table

	trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type	shape_dist_traveled
1	1170717	1970-01-01 20:24:47.000	1970-01-01 20:24:47.000	10874	16		0	0	6.6001
2	1170717	1970-01-01 20:25:52.000	1970-01-01 20:25:52.000	14369	17		0	0	6.8933
3	1170717	1970-01-01 20:27:03.000	1970-01-01 20:27:03.000	10875	18		0	0	7.2144
4	1170717	1970-01-01 20:28:00.000	1970-01-01 20:28:00.000	10876	19		0	0	7.4454
5	1170717	1970-01-01 20:29:04.000	1970-01-01 20:29:04.000	10877	20		0	0	7.9465
6	1170717	1970-01-01 20:29:26.000	1970-01-01 20:29:26.000	10878	21		0	0	8.1205
7	1170717	1970-01-01 20:29:47.000	1970-01-01 20:29:47.000	10879	22		0	0	8.2915
8	1170717	1970-01-01 20:30:21.000	1970-01-01 20:30:21.000	10880	23		0	0	8.5626
9	1170717	1970-01-01 20:31:15.000	1970-01-01 20:31:15.000	10881	24		0	0	8.9859
10	1170717	1970-01-01 20:31:39.000	1970-01-01 20:31:39.000	10882	25		0	0	9.1742
11	1170717	1970-01-01 20:32:12.000	1970-01-01 20:32:12.000	10883	26		0	0	9.4336
12	1170717	1970-01-01 20:32:37.000	1970-01-01 20:32:37.000	10884	27		0	0	9.6342
13	1170717	1970-01-01 20:33:05.000	1970-01-01 20:33:05.000	10885	28		0	0	9.8569
14	1170717	1970-01-01 20:33:35.000	1970-01-01 20:33:35.000	10886	29		0	0	10.0938
15	1170717	1970-01-01 20:33:56.000	1970-01-01 20:33:56.000	10887	30		0	0	10.2582
16	1170717	1970-01-01 20:35:11.000	1970-01-01 20:35:11.000	14415	31		0	0	10.8446
17	1170717	1970-01-01 20:35:37.000	1970-01-01 20:35:37.000	14957	32		0	0	11.0547

(c) Stop & time table

	stop_id	stop_code	stop_name	stop_desc	stop_lat	stop_lon	zone_id	stop_url	location_type	parent_station	wheelchair_boarding
1	3	NULL	Park/Speedway	N PARK AVE & E SPEEDWAY BLVD	32.236753	-110.956685	NULL	NULL	NULL	NULL	0
2	4	NULL	Glenn/Campbell	E GLENN ST & N CAMPBELL AV	32.257543	-110.943543	NULL	NULL	NULL	NULL	0
3	5	NULL	Glenn/Alvernon	E GLENN ST & N ALVERNON WAY	32.257785	-110.909233	NULL	NULL	NULL	NULL	0
4	8	NULL	Glenn/Alvernon	E GLENN ST & N HASKELL DR	32.257945	-110.910127	NULL	NULL	NULL	NULL	0
5	9	NULL	Glenn/Campbell	E GLENN ST & N CAMPBELL AV	32.257615	-110.943542	NULL	NULL	NULL	NULL	0
6	14	NULL	Forgeus/36th St	S FORGEUS STRAV & E 36TH ST	32.192682	-110.935772	NULL	NULL	NULL	NULL	0
7	18	NULL	UAMC/South Campus	UAMC KINO	32.17689	-110.930741	NULL	NULL	NULL	NULL	0
8	20	NULL	Escalante/Pantano	E ESCALANTE RD & S PANTANO RD	32.17723	-110.823121	NULL	NULL	NULL	NULL	0
9	21	NULL	Stella/Kolb	E STELLA RD & S KOLB RD	32.184746	-110.841452	NULL	NULL	NULL	NULL	0
10	22	NULL	Wilmot/Broadway	S WILMOT RD & E BROADWAY BLVD	32.220401	-110.8581	NULL	NULL	NULL	NULL	0
11	23	NULL	5th St/Alvernon	E 5TH ST & N ALVERNON WAY	32.228973	-110.91003	NULL	NULL	NULL	NULL	0
12	25	NULL	Downtown Ronstadt Center	W RONSTADT TRANSIT CENTER DR	32.222563	-110.968047	NULL	NULL	NULL	NULL	0
13	28	NULL	Downtown Ronstadt Center	W RONSTADT TRANSIT CENTER DR	32.222511	-110.968283	NULL	NULL	NULL	NULL	0
14	29	NULL	6th St/Campbell	E 6TH ST & N CAMPBELL AVE	32.227727	-110.942714	NULL	NULL	NULL	NULL	0
15	30	NULL	5th St/Alvernon	E 5TH ST & N ALVERNON WAY	32.228805	-110.909321	NULL	NULL	NULL	NULL	0
16	32	NULL	Stella/Kolb	E STELLA RD & S KOLB RD	32.184594	-110.840733	NULL	NULL	NULL	NULL	0
17	34	NULL	Pima College East Campus	S FRED ENKE DR & E PCC EAST PL	32.166109	-110.815521	NULL	NULL	NULL	NULL	0
18	35	NULL	Harrison/Golf Links	S HARRISON RD & E GOLF LINKS RD	32.192592	-110.789623	NULL	NULL	NULL	NULL	0
19	36	NULL	Speedway/Harrison Park ...	E SH PARK N RIDE WAY & N EUCLI...	32.23501	-110.790247	NULL	NULL	NULL	NULL	0
20	40	NULL	Speedway/Alvernon	E SPEEDWAY BLVD & N ALVERNO...	32.236315	-110.910143	NULL	NULL	NULL	NULL	0

(d) Stops

	route_id	service_id	trip_id	trip_headsign	trip_short_name	direction_id	block_id	shape_id
1	11617	2	1168920	TOHONO CENTER	NULL	0	381125	35916
2	11617	2	1168921	TOHONO CENTER	NULL	0	381126	35916
3	11617	2	1168922	TOHONO CENTER	NULL	0	381127	35916
4	11617	2	1168923	TOHONO CENTER	NULL	0	381125	35916
5	11617	2	1168924	TOHONO CENTER	NULL	0	381126	35916
6	11617	2	1168925	TOHONO CENTER	NULL	0	381127	35916
7	11617	2	1168926	TOHONO CENTER	NULL	0	381125	35916
8	11617	2	1168927	TOHONO CENTER	NULL	0	381126	35916
9	11617	2	1168928	TOHONO CENTER	NULL	0	381127	35916
10	11617	2	1168929	TOHONO CENTER	NULL	0	381125	35916
11	11617	2	1168930	TOHONO CENTER	NULL	0	381126	35916
12	11617	2	1168931	TOHONO CENTER	NULL	0	381127	35916
13	11617	2	1168932	TOHONO CENTER	NULL	0	381125	35916
14	11617	2	1168933	TOHONO CENTER	NULL	0	381126	35916
15	11617	2	1168947	DOWNTOWN	NULL	1	381125	35917
16	11617	2	1168934	DOWNTOWN	NULL	1	381126	35917
17	11617	2	1168935	DOWNTOWN	NULL	1	381127	35917
18	11617	2	1168936	DOWNTOWN	NULL	1	381125	35917
19	11617	2	1168937	DOWNTOWN	NULL	1	381126	35917
20	11617	2	1168938	DOWNTOWN	NULL	1	381127	35917
21	11617	2	1168939	DOWNTOWN	NULL	1	381125	35917

(e) Trips

FIGURE 5 Parsed GTFS static data in database

	header.timestamp	tripupdate.timestamp	trip.schedule	trip.trip_id	trip.route_id	trip.start_time	trip.start_date	vehicle.id	vehicle.label	vehicle.license_plate	stop_sequence	stop_id	arriv
1	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1152912	16	23:10:00	20140630	-1	2518	-1	39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,...	ORFS(NW),ORPL(SW),ORLE(SW),DR11(SW),DR9A(SE),STS...	0,0,
2	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1154874	8	23:04:00	20140630	-1	2525	-1	49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,...	BRMA(SE),BRJE(SE),BRLA(SE),BRKO(SE),BRMO(SE),BRPR(...	0,0,
3	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1154965	8	23:18:00	20140630	-1	2536	-1	50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,...	RTC(08S),6A12(SW),6A15(SW),6A18(SW),6A20(NW),6A22(S...	0,0,
4	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1152836	16	22:48:00	20140630	-1	3011	-1	61,62,63,64,65,66,67,68,	IN2300WB,INCC(NW),INSW(NE),INSH(NW),IN3220WB,INTI(...	0,0,
5	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1154964	8	22:48:00	20140630	-1	3101	-1	58,59,60,61,62,63,64,65,66,67,68,	6A34(SW),6A37(NW),6A39(SW),6A44(NW),6AVE(SW),6AAJ(S...	0,0,
6	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1154022	4	23:30:00	20140630	-1	3109	-1	21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,...	SPAT(SE),SPCO(SE),SPBL(SE),SPSW(SE),SPAR(SE),SPRO(...	0,0,
7	2014-07-01 00:00:05.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
8	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1152912	16	23:10:00	20140630	-1	2518	-1	39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,...	ORFS(NW),ORPL(SW),ORLE(SW),DR11(SW),DR9A(SE),STS...	0,0,
9	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1154874	8	23:04:00	20140630	-1	2525	-1	49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,...	BRMA(SE),BRJE(SE),BRLA(SE),BRKO(SE),BRMO(SE),BRPR(...	0,0,
10	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1152836	16	22:48:00	20140630	-1	3011	-1	61,62,63,64,65,66,67,68,	IN2300WB,INCC(NW),INSW(NE),INSH(NW),IN3220WB,INTI(...	0,0,
11	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
12	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1152836	16	22:48:00	20140630	-1	3011	-1	61,62,63,64,65,66,67,68,	IN2300WB,INCC(NW),INSW(NE),INSH(NW),IN3220WB,INTI(...	0,0,
13	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1154965	8	23:18:00	20140630	-1	2536	-1	51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,...	6A12(SW),6A15(SW),6A18(SW),6A20(NW),6A22(SW),6A27(S...	0,0,
14	2014-07-01 00:00:35.000	1969-12-31 17:00:00.000	0	1154022	4	23:30:00	20140630	-1	3109	-1	21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,...	SPAT(SE),SPCO(SE),SPBL(SE),SPSW(SE),SPAR(SE),SPRO(...	0,0,
15	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1152912	16	23:10:00	20140630	-1	2518	-1	39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,...	ORFS(NW),ORPL(SW),ORLE(SW),DR11(SW),DR9A(SE),STS...	0,0,
16	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1154874	8	23:04:00	20140630	-1	2525	-1	49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,...	BRMA(SE),BRJE(SE),BRLA(SE),BRKO(SE),BRMO(SE),BRPR(...	0,0,
17	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1154965	8	23:18:00	20140630	-1	2536	-1	51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,...	6A12(SW),6A15(SW),6A18(SW),6A20(NW),6A22(SW),6A27(S...	0,0,
18	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1154022	4	23:30:00	20140630	-1	3109	-1	21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,...	SPAT(SE),SPCO(SE),SPBL(SE),SPSW(SE),SPAR(SE),SPRO(...	0,0,
19	2014-07-01 00:01:05.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
20	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1152912	16	23:10:00	20140630	-1	2518	-1	47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,...	RTC(16S),ALST(NE),CHPE(NW),CHOC(NW),MASI(NW),MA17...	0,0,
21	2014-07-01 00:02:35.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
22	2014-07-01 00:03:05.000	1969-12-31 17:00:00.000	0	1152836	16	22:48:00	20140630	-1	3011	-1	61,62,63,64,65,66,67,68,	IN2300WB,INCC(NW),INSW(NE),INSH(NW),IN3220WB,INTI(...	0,0,
23	2014-07-01 00:03:05.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
24	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1154874	8	23:04:00	20140630	-1	2525	-1	49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,...	BRMA(SE),BRJE(SE),BRLA(SE),BRKO(SE),BRMO(SE),BRPR(...	0,0,
25	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1154965	8	23:18:00	20140630	-1	2536	-1	51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,...	6A12(SW),6A15(SW),6A18(SW),6A20(NW),6A22(SW),6A27(S...	0,0,
26	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1152836	16	22:48:00	20140630	-1	3011	-1	61,62,63,64,65,66,67,68,	IN2300WB,INCC(NW),INSW(NE),INSH(NW),IN3220WB,INTI(...	0,0,
27	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1154022	4	23:30:00	20140630	-1	3109	-1	21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,...	SPAT(SE),SPCO(SE),SPBL(SE),SPSW(SE),SPAR(SE),SPRO(...	0,0,
28	2014-07-01 00:01:35.000	1969-12-31 17:00:00.000	0	1153770	3	22:39:00	20140630	-1	3128	-1	73,74,75,76,	PADO(SW),PAPO(SW),PACN(SW),PCCE(ES),	0,0,
29	2014-07-01 00:02:05.000	1969-12-31 17:00:00.000	0	1152912	16	23:10:00	20140630	-1	2518	-1	47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,...	RTC(16S),ALST(NE),CHPE(NW),CHOC(NW),MASI(NW),MA17...	0,0,

(a) Trip update table

	header.timestamp	trip.schedule	trip.trip_id	trip.route_id	trip.start_time	trip.start_date	position.latitude	position.longitude	position.bearing	position.odometer	position.speed
1	2014-07-01 00:00:05.000	0	1152912	-1	-1	-1	32.222984314	-110.968032837	0	0	0
2	2014-07-01 00:00:05.000	0	1154874	-1	-1	-1	32.2208518982	-110.840194702	0	0	0
3	2014-07-01 00:00:05.000	0	1154965	-1	-1	-1	32.2229652405	-110.968208313	0	0	0
4	2014-07-01 00:00:05.000	0	1152836	-1	-1	-1	32.3375854492	-111.044448853	0	0	0
5	2014-07-01 00:00:05.000	0	1154964	-1	-1	-1	32.2724342346	-111.007247925	0	0	0
6	2014-07-01 00:00:05.000	0	1154022	-1	-1	-1	32.2360877991	-110.876296997	0	0	0
7	2014-07-01 00:00:05.000	0	1153770	-1	-1	-1	32.1568336487	-110.907775879	0	0	0
8	2014-07-01 00:00:35.000	0	1152912	-1	-1	-1	32.222984314	-110.968032837	0	0	0
9	2014-07-01 00:00:35.000	0	1154874	-1	-1	-1	32.2208518982	-110.840194702	0	0	0
10	2014-07-01 00:00:35.000	0	1154965	-1	-1	-1	32.2234954834	-110.968330383	0	0	0
11	2014-07-01 00:00:35.000	0	1152836	-1	-1	-1	32.3375854492	-111.044448853	0	0	0
12	2014-07-01 00:00:35.000	0	1154022	-1	-1	-1	32.2360725403	-110.875244141	0	0	0
13	2014-07-01 00:00:35.000	0	1153770	-1	-1	-1	32.1568336487	-110.907775879	0	0	0
14	2014-07-01 00:01:05.000	0	1152912	-1	-1	-1	32.2227973938	-110.968215942	0	0	0
15	2014-07-01 00:01:05.000	0	1154874	-1	-1	-1	32.2208404541	-110.831634521	0	0	0
16	2014-07-01 00:01:05.000	0	1154965	-1	-1	-1	32.2234954834	-110.968330383	0	0	0
17	2014-07-01 00:01:05.000	0	1152836	-1	-1	-1	32.3375778198	-111.046508789	0	0	0
18	2014-07-01 00:01:05.000	0	1154022	-1	-1	-1	32.2360725403	-110.875244141	0	0	0
19	2014-07-01 00:01:05.000	0	1153770	-1	-1	-1	32.1640930176	-110.920669556	0	0	0
20	2014-07-01 00:01:35.000	0	1152912	-1	-1	-1	32.2228775024	-110.968238831	0	0	0
21	2014-07-01 00:01:35.000	0	1154874	-1	-1	-1	32.2208404541	-110.831634521	0	0	0
22	2014-07-01 00:01:35.000	0	1154965	-1	-1	-1	32.2190704346	-110.968635559	0	0	0
23	2014-07-01 00:01:35.000	0	1152836	-1	-1	-1	32.3375778198	-111.046508789	0	0	0
24	2014-07-01 00:01:35.000	0	1154022	-1	-1	-1	32.2360725403	-110.870368958	0	0	0
25	2014-07-01 00:01:35.000	0	1153770	-1	-1	-1	32.1640930176	-110.920669556	0	0	0
26	2014-07-01 00:02:05.000	0	1152912	-1	-1	-1	32.2228775024	-110.968238831	0	0	0

(b) Vehicle position table

FIGURE 6 Parsed GTFS real-time data in database

4.3 Network Construction

Multiple transport modes were considered in our intercity passenger traveling problem. To model this multimodal transport network and optimize route selection according to passenger preferences, it was helpful to build the network in an efficient way. Different network-building methods have been studied since multimodal networks can be viewed from many aspects [19]. Networks can be viewed as physical systems and classified as road, rail, water, and air, or they can be viewed as functional systems, classified into private modes (e.g. foot, bike and car) and public modes (e.g. bus, train, metro). Two major network structures have been used extensively in transportation planning: node-based [20] and link-based [21] networks. Recently, trip-based network representation was proposed in [22], in this case transit vehicle trips are used as network edges and transfer stop hierarchy is taken into account. Another lane-based network presented in [23] could serve as a more realistic platform to provide a geospatial context for traffic simulations to be performed at the level of individual vehicles. A new super-network platform was constructed [24], which was an expanded network in which activity links were introduced into the conventional time-space network.

In our multimodal network, transit buses, walking, and flights were considered within two cities, Boston and Tucson. In order to fit and make better use of the GTFS data format in our SQL database (see Section 4.1), individual transit bus networks were established for the two cities initially, and then these two networks were flattened to integrate walking and transit, and finally the two networks were connected by the flight network. This network building process was accomplished using the package 'NetworkX' developed for the Python programming language [25]. Database interaction between SQL and Python can be found in [26] using package 'pymssql'. The example Python code is shown below: 'host' represents the IP address of the database server, '1433' is the default value of the connection port, 'database' is the name used for the database in database management system.

```
SQL = pymssql.connect(
        host='xxx.xxx.xxx.xxx',
        port=1433,
        user='xxxx',
        password='xxxx',
        database='xxxx'
)

SQL.cursor()
```

Figure 7 shows detailed flow diagram of Python + SQL to generate the static multimodal transport network with corresponding static attributes (taking Tucson for example). Figure 8 shows a query example in SQL database server.

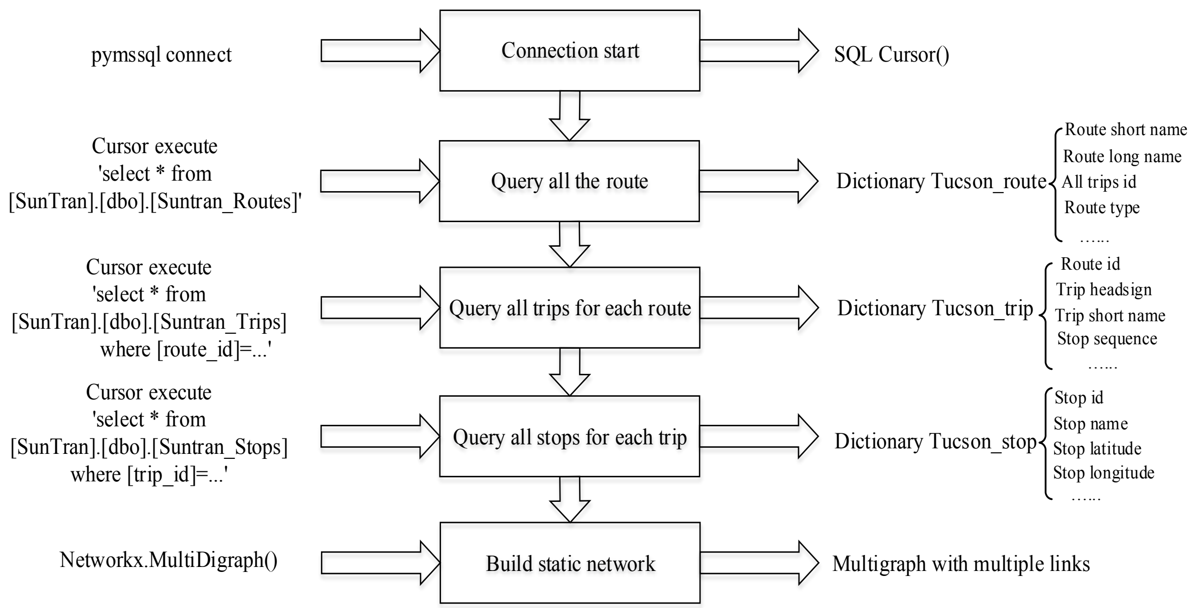


FIGURE 7 Flow diagram of Python + SQL to generate static multimodal network

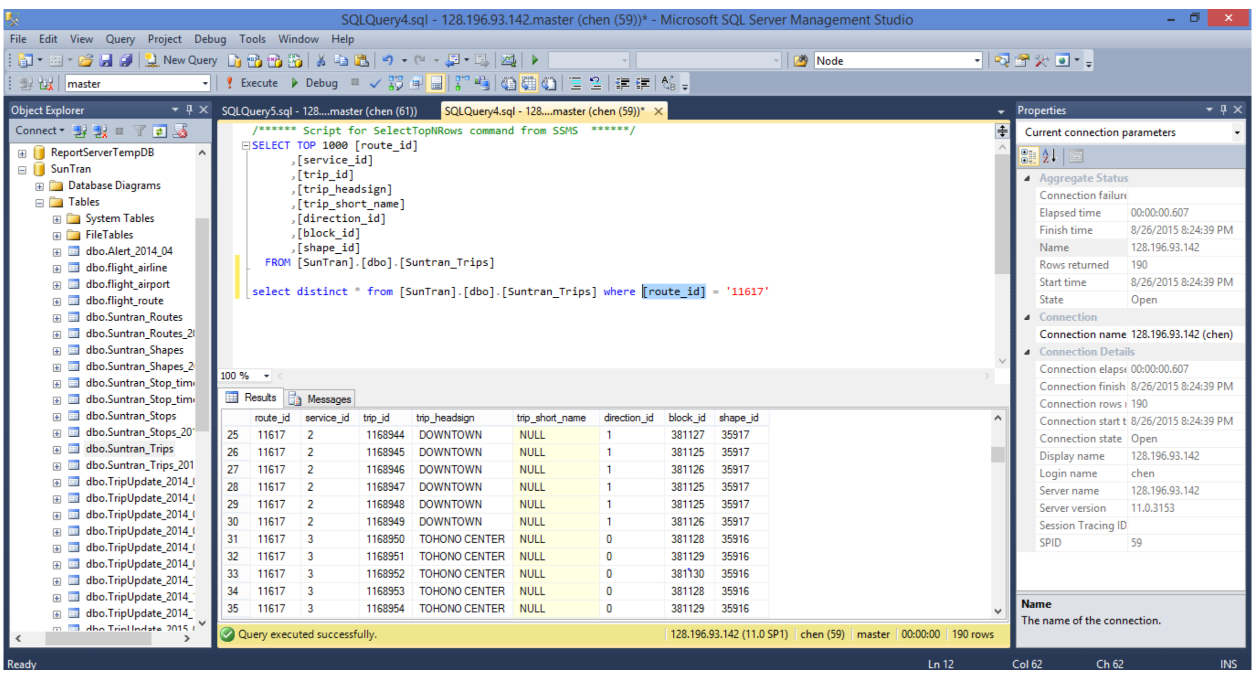


FIGURE 8 Example in SQL to query all the information for route 11617

After separate transport networks for the two cities were built, the overall intercity network was connected via airports and constrained by the flight network. Land-based travel modes between the two cities were considered impractical for most travelers. For a practical trip between Tucson and Boston, the most unalterable constraint is the flight network, which has a strong time-property. Therefore, arriving with high reliability well before the flight takeoff time, which would be a parameter input by the user, was critical in trip planning.

It should be noted that for this established directed graph (Figure 9), only static attributes were assigned for nodes (e.g. identification number, name, coordinates, scheduled departure, and arrival time) and links (e.g. parent nodes, succeeding nodes, and scheduled travel time for this link). However, for the real-time optimization, real-time information about traffic interruption, possible delay, and all other unexpected alterations needed to be accessible. Accordingly, partial or entire networks needed to be updated once the real-time feedback information had been imported into the database. This updating process was completed by creating event alerts in the database to update automatically.

Figure 9 demonstrates a simple transit network consisting of three routes and 13 stops. Multiple routes may travel on the same link and pass through the same stops. The link travel time and link travel time reliability are estimated by specific routes and links. For example, westbound routes A and C are designed to travel on a link consisting of Stops 1 and 2. $TT_{A,1}^{tod,w}$ and $TTR_{A,1}^{tod,w}$ represent the travel time and travel time reliability on the first link, route A, given a TOD and w, respectively. $T_{C,1}^{tod,w}$ and $TTR_{C,1}^{tod,w}$ represent the travel time and corresponding travel time reliability on the same link, route C. The transit stop information was extracted from the GTFS static data and the links were constructed using two consecutive stops on a specific route. The link travel times and link travel time reliability were estimated using the GTFS real-time data. Then, a network was completed using the transit stops, constructed links, and corresponding link travel time and travel time reliability.

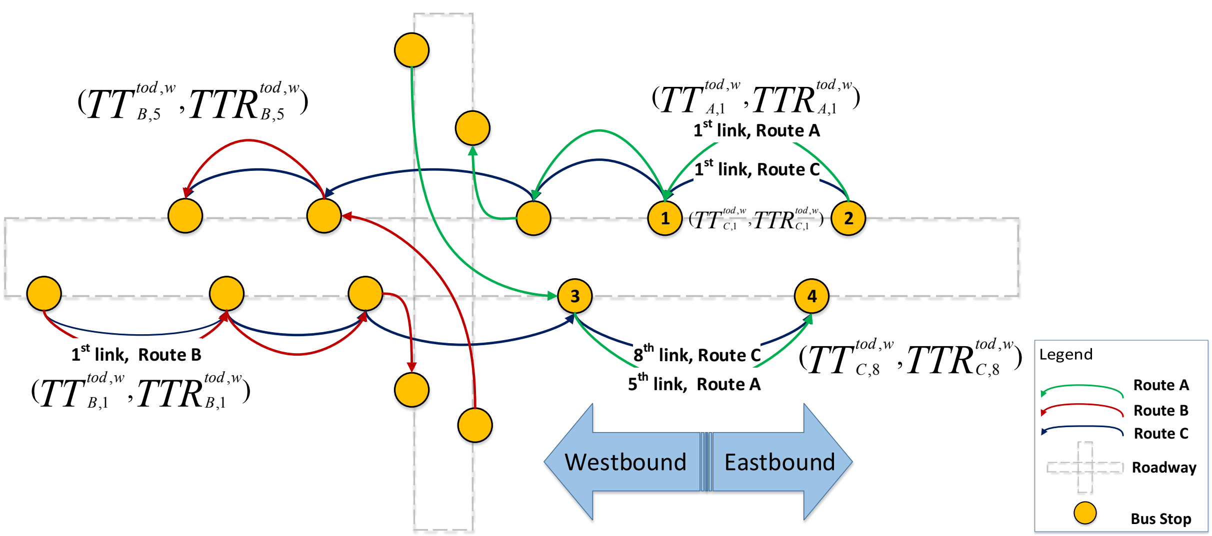


FIGURE 9 Demonstration of transit network construction

4.4Chance Constrained Decision Model

Due to the complexity and randomness of real-world situations, it is difficult to determine travel time explicitly in a large-scale transit network. Additionally, walking is always involved in public transit and more complexity is added when taking into account the uncertainty of transfer time and waiting time. For our study, the airport was set as the destination considering that travelers are concerned about on-time arrival confidence level because of the significant inconvenience of missing a flight. Finding an optimal path to guarantee arriving the airport on time at high likelihood was a priori preference. Practically, in the stochastic transit network, the route with lowest travel time might not be the best choice for passengers to catch the flight during specific hours on specific day.

For this kind of problem, chance constrained programming was developed as a means of describing constraints in mathematical programming models in the form of probability levels of attainment. Consideration of chance constraints allows decision makers to consider mathematical programming objective in terms of the probability of their attainment [27]. In this research, the uncertainties in travel time were considered and a deterministic model was extended to be a stochastic model with some chance constraints to guarantee the probability of on-time arrival. For clarity, the travel time between two adjacent stops by bus is termed as link travel time while travel time from the origin to the destination is referred to as path travel time. Accordingly, link travel time is the element for path travel time.

Considering a stochastic and time-dependent network, total travel time is the objective function and mainly consists of link travel time, transfer time (same as waiting time), and walking time. However, for link travel time, transportation literature does not provide a universally valid model for bus movements in an urban environment since they are strongly affected by vehicular and passenger traffic conditions, road organization, traffic signal control management, company policies, etc. [28]. Therefore, using nonparametric probability distribution estimation methods could provide greater flexibility and increased fidelity with fewer assumptions. Details about estimated travel time distribution can be found in our previous research [18]. Specific calculation of mean value of link travel time is shown in the following data preparation section. All corresponding assumptions made for this study and the notations (Table 1) used for the model are summarized below.

1. The modes studied here are transit bus and walking, and transfer mode is used as dummy mode for modeling convenience. Walking was not considered at original stops, which means walking was not chosen at the departure stop.
2. Link travel time, transfer time, and walking time between any consecutive nodes were all treated as random variables. The dependency was not considered.
3. Bus transfer time was assumed to follow uniform distribution, and the lower and upper limit were determined by scheduled bus timetable.
4. Walking time was computed based on distance between two nodes and walking speed which was assumed to follow a normal distribution $N(\mu, \sigma^2)$. Reasonable values of μ and σ are 1.35m/s and 0.2m/s [29]. Since the first and second moment of reciprocal normal distribution does not exist, the mean of walking time was estimated by $distance/\mu$, and the standard deviation of walking time was estimated by the estimator $distance * (\frac{1}{\mu-0.67449*\sigma} - \frac{1}{\mu+0.67449*\sigma})/1.34898$ [30].

TABLE 1 All notations used in the model

Notations	Description
N	Set of nodes in the transit network, with index i, j
O	Departure node
D	Destination node
A	Set of links in the multi-modal network
M	Set of travel modes in the multi-modal network, with index m
$E_{t,ijm}$	Mean value of travel time for link (i, j) under mode m
$SD_{t,ijm}$	Standard deviation of travel time for link (i, j) under mode m
AT	User-defined anticipated arrival time
CT	Current time
c	User-defined on-time arrival confidence level
z_r	Quantile of normal distribution at confidence level c
x_{ijm}	Binary variable, selecting mode m for link (i, j) or not

For this stochastic network, the objective of minimizing expected total travel time can be expressed as

$$min\, ET = \sum_{(i,j)\in A, m\in M} x_{ijm} E_{t,ijm} \tag{3}$$

For the constraints, the basic flow balance constraints should be included to generate feasible path which is given below.

$$\sum_{(i,j)\in A,m\in M} x_{ijm} - \sum_{(j,i)\in A,m\in M} x_{ji} = \begin{cases} 1, & if\ i = O \\ 0, if\ i \neq O\ or\ D \\ -1, & if\ i = D \end{cases} \tag{4}$$

The chance constraint is introduced here to guarantee the on-time arrival probability which should be greater than pre-defined confidence level c.

$$P\{ET \leq (AT - CT)\} \geq c \tag{5}$$

Travel time uncertainties are typically represented by random distributions. If the travel time between any consecutive stops are independently distributed, the path travel time follows a normal distribution approximately following the central limit theorem. Based on the central limit theorem and independence assumption, the mean and variance values of all the arcs can be added together as mean and variance value of the path. Hence, the chance constraint can be formed in the following equivalent deterministic constraints according to [31].

$$ET \leq \sum_{(i,j)\in A,m\in M} x_{ijm} E_{t,ijm} - z_r \sqrt{\sum\nolimits_{(i,j)\in A,m\in M} x_{ijm} (SD_{t,ijm})^2} \tag{6}$$

$$x_{ijm} \in \{0,1\} \tag{7}$$

Here, different random distributions could be included in this model, also it is more easily implemented based on existing efficient shortest path algorithms.

4.5Solution Method

Up to this point, the model with chance constraints had been transformed to a classic network model which is similar to a shortest path problem with an extra travel time upper limit constraint. Numerous algorithms have been developed for this problem category in static and stochastic networks. Classic shortest path algorithms such as Dijkstra, Bellman, and Dreyfus focus on networks with deterministic arc weights. For time-dependent networks, algorithms including exact or heuristic algorithms were also proposed recently [32,33]. All of these algorithms seek to obtain an optimum or near optimum path which limits alternative options.

For practical applications, some paths have the same objective value and satisfy the confidence level requirement, therefore several possible optimal options should be ranked for passengers to choose from based on their own preferences. Obtaining several paths in increasing order of length is usually referred as the *k*-shortest paths problem which is a generalization of the shortest path problem [34]. The *k*-shortest path problem was originally examined by Hoffman and Pavley [35], but nearly all early attempts to solve it led to exponential time algorithms [34]. The best known implementation for this algorithm was proposed by Yen [36] using modern data structures, in which $O(kn(m + nlog(n)))$ limits the worst case complexity. This algorithm essentially performs $O(n)$ single-source shortest path computations for each output path. Based on these considerations, the framework of the k-shortest path algorithm was used to conduct our experiments. In the repeated k iterations of the algorithm, the feasibility of constraints [4] needed to be checked; the solution path would not be stored if the feasibility was not satisfied.

4.6 User Interface Prototype

The graphic user interface (GUI) in this section was developed based on the Python + Eclipse environment. Several GUI packages were available in the Python library, for instance, TkInter, Traits/TraitsUI, and gui2py. For this project, WxPython was used [37]. Tucson and Boston were considered since their public transit data formats were very similar and easily integrated together.

For intercity travel from Tucson to Boston, travelers begin with a departure bus stop in Tucson (generally near their home) and terminate at a bus stop in Boston (generally near their final destination). Normally, flight tickets will be booked well in advance of the beginning of their trip. There might be some connections in their flight, however, the departure time from home and arrive time at destination are constrained by the takeoff time of their flight from Tucson and landing time in Boston.

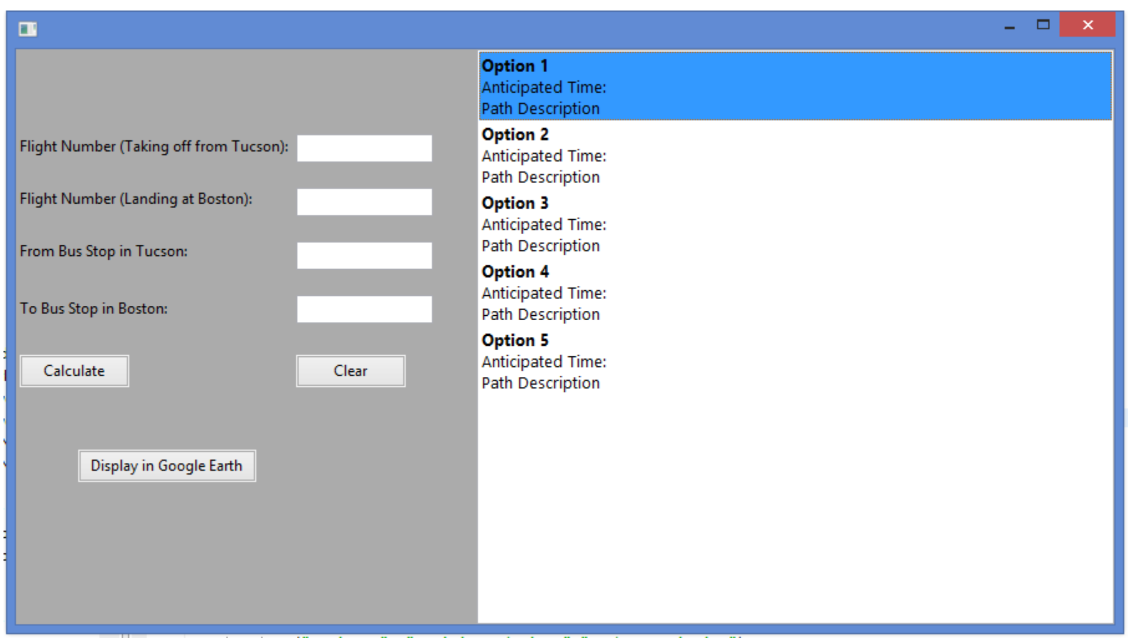


FIGURE 10 Prototype of graphic user interface for intercity trip support

Therefore, four inputs were gathered from users: two flight numbers (taking off from Tucson and landing at Boston) and two bus stops, as shown in Figure 10. The 'Calculate' button was used to obtain at most the five best options based on anticipated travel time. 'Display in Google Earth' created a kml file [38,39] of the chosen path allowing users to visualize it in Google Earth. The 'Clear' button cleared all the input and path results.

5. DISCUSSION OF RESULTS

5.1System Demonstration

<u>Case Study I</u>: Suppose Delta Airlines has been chosen by the traveler, with the flight information shown in Figure 11. Flight DL 1345 takes off at 6:15am (not a peak hour) from the Tucson International Airport and Flight DL 1500 will land at 4:42pm (around peak hour) at Boston Logan International Airport, with a connection in Atlanta.

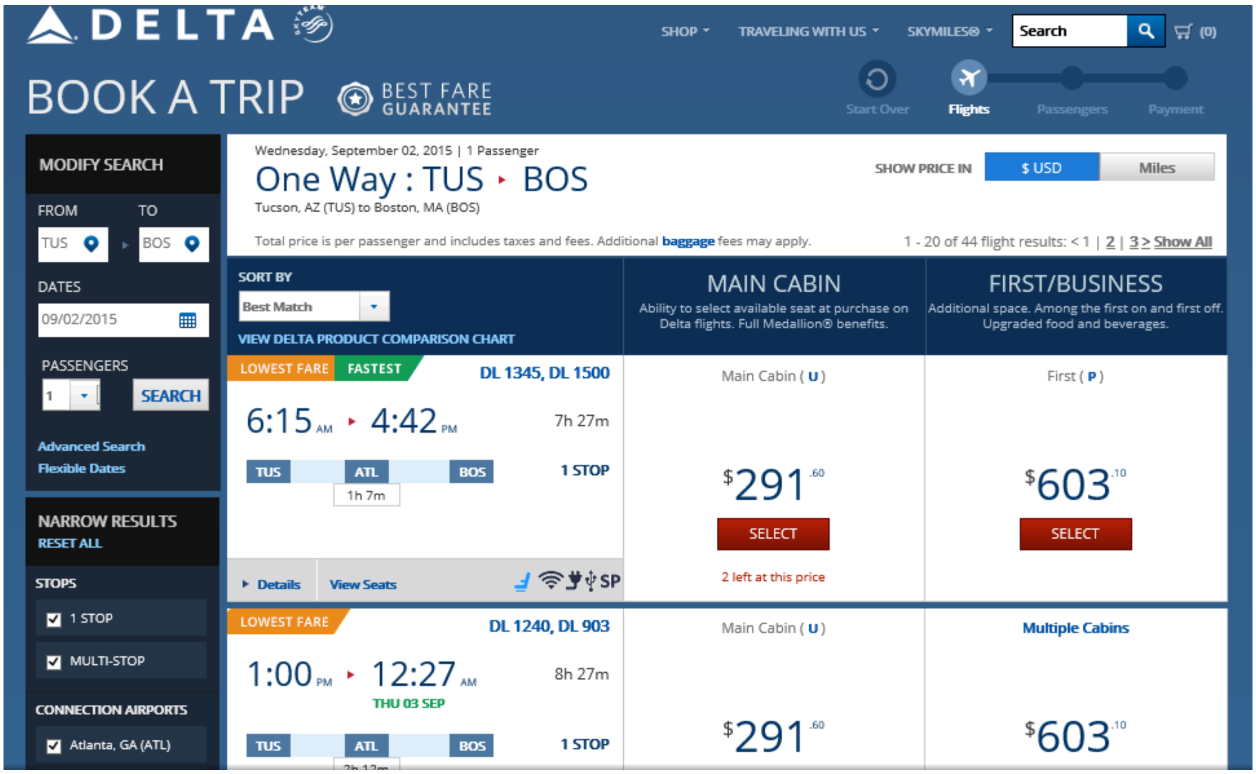


FIGURE 11 Flight information from Delta Airline's website

After inputting the flight information in the user interface, the recommended path and corresponding travel time is given in Figure 12. An additional time of 25 minutes would be needed to guarantee on-time arrive at Tucson International Airport if the traveler boarded the flight at 6 am. After the path results are calculated, Google Earth can be used to display the results. 'Tucson, Option 1' and 'Boston, Option 1' are displayed respectively in Figure 13 and Figure 14. The marker 'S' and 'D' in Figure 13 and Figure 14 represent the start stop and ending stops. Walking mode is indicated by two 'W's.

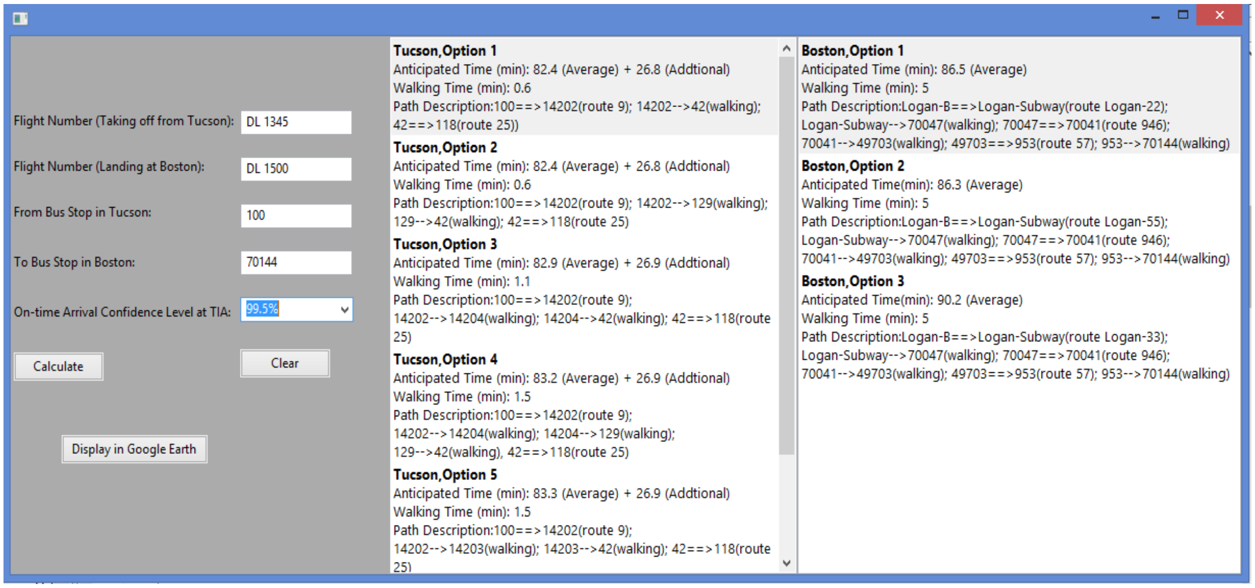


FIGURE 12   Optimized results for case I

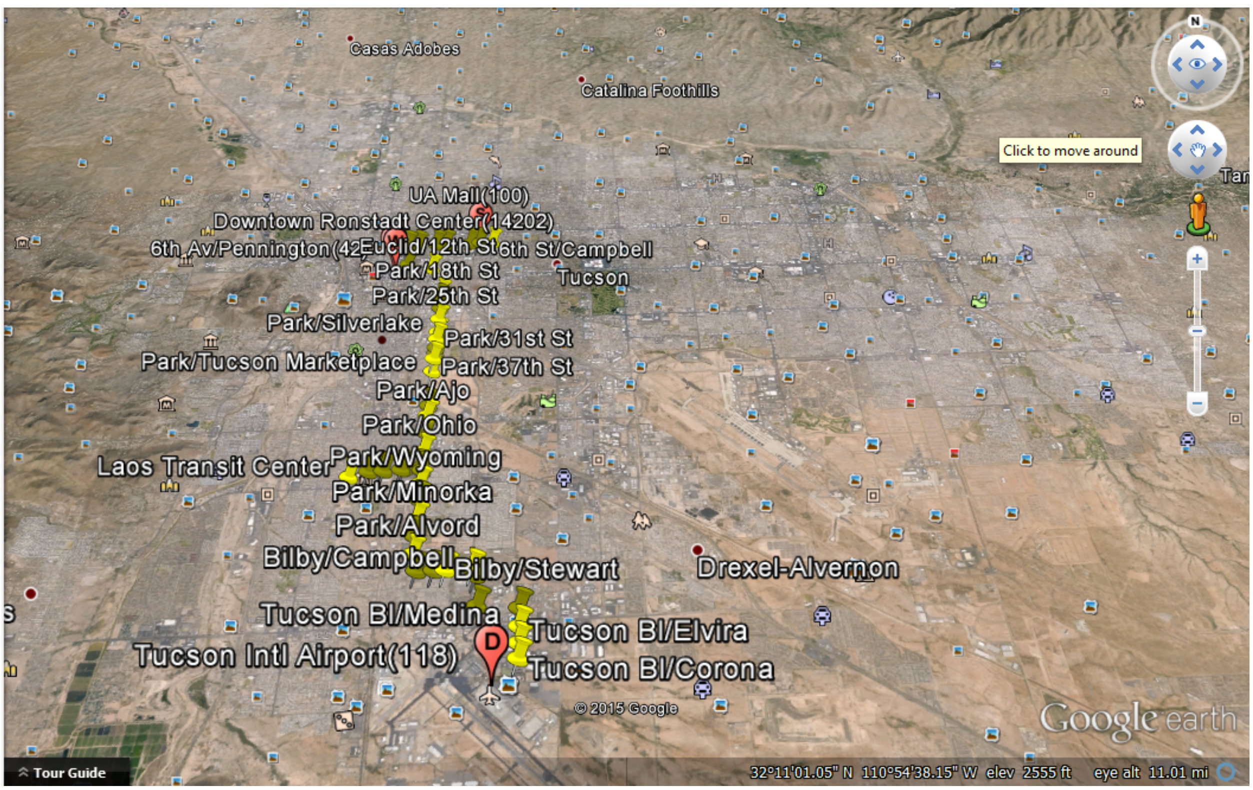


FIGURE 13   Path displayed in Google Earth for case I, Tucson

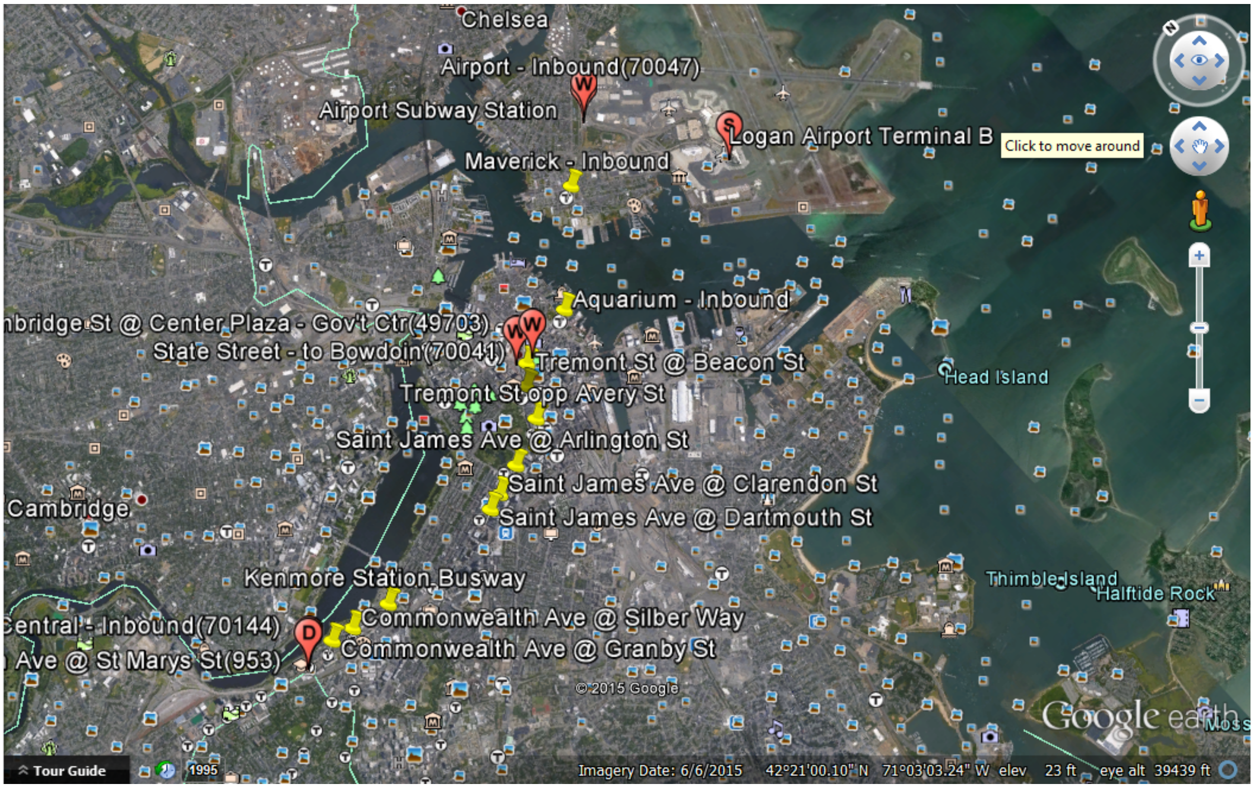


FIGURE 14 Path displayed in Google Earth for Case I, Boston

Case Study II: In case II, American Airlines is chosen (Figure 15), and the takeoff time is around the traffic peak hour of 5 pm and landing time is 10:29 am the next day in Boston. The optimized results and displayed path are shown in Figure 16 and Figure 17.

SELECT — Average Price per Person - 621.20 USD
Price and Tax Information

Fare Alert: Nonrefundable fare. Change fee associated with fare.

Flight Alert: Return airport different than originating airport.

Carrier	Flight #	Departing		Arriving		Aircraft Type	Cabin	Flight Miles	Meals	Travel Time
		City	Date & Time	City	Date & Time					
AMERICAN AIRLINES OPERATED BY SKYWEST AIRLINES AS AMERICAN EAGLE	2888	TUS Tucson	Sep 08, 2015 05:37 PM	LAX Los Angeles	Sep 08, 2015 07:15 PM	CRJ	Economy View Seats	451	N/A	1 hr 38 min
AMERICAN AIRLINES OPERATED BY US AIRWAYS	650	LAX Los Angeles	Sep 09, 2015 10:10 PM	PHL Philadelphia	Sep 10, 2015 06:29 AM	321	Economy		Food For Purchase	5 hr 19 min
Alert: Overnight flight or connection.										
AMERICAN AIRLINES OPERATED BY US AIRWAYS	1863	PHL Philadelphia	Sep 10, 2015 09:15 AM	BOS Boston	Sep 10, 2015 10:29 AM	E90	Economy		N/A	1 hr 14 min

FIGURE 15 Flight information from American Airline's website

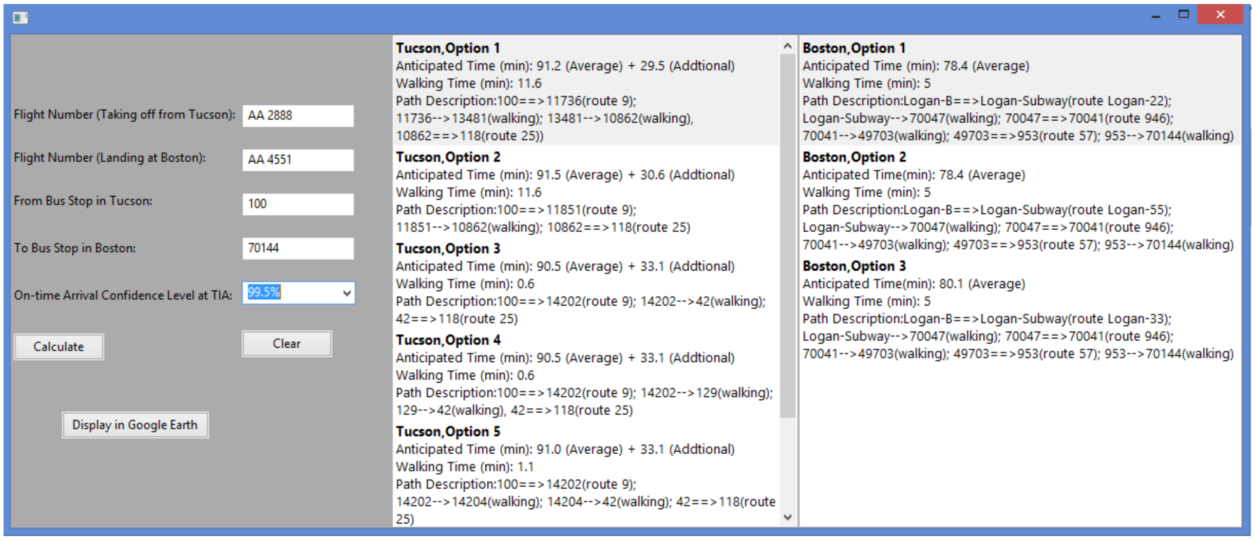


FIGURE 16 Optimized results for case II

FIGURE 17 Path displayed in Google Earth for case II, Tucson

5.2Sensitivity Analysis

Based on the abovementioned data and the approaches for network construction, the transit network in the Tucson area was constructed for planning travelers' trips. The total transit bus network consisted of approximately 2,332 bus stops and 3,529 links. Two modes were primarily considered in the network: walking and transit buses. Three groups of experiments were created to demonstrate the effects of transit service uncertainty on path choices. These three groups of experiments focused on the local portion of the Tucson to Boston trip, with the Tucson International Airport as the trip destination.

<u>Experiment I - Effects of Chance Constraint on Path Choice</u>

This group of experiments was designed to investigate the effects of a chance constraint on path choice. The departure time was selected as 5 pm on a weekday, when traffic usually suffered from recurrent congestion. The confidence level of the chance constraint was set at 99.5%. Four scenarios were created by selecting two origins (including the University of Arizona mall bus stop (Stop 100) and Kain/Kimberly PI bus stop (Stop 13912)) and whether or not the chance constraint was considered.

- Scenario 1a; origin: Stop 100; do not consider chance constraint;
- Scenario 1b; origin: Stop 100; consider chance constraint;
- Scenario 2a; origin: Stop 13912; do not consider chance constraint;
- Scenario 2b; origin: Stop 13912; consider chance constraint;

The details of the results are listed in Table 2, and several findings are summarized below.

TABLE 2 Results comparisons between with and without chance constraints

Departure Stop	Consider chance constraint	Optimal travel time (minutes)	Choose walking	Total walk time (minutes)	Optimal path
100	No	90.55	Once	0.6	$100 \xrightarrow{Route\ 9} 14202 \xrightarrow{Walking} 42 \xrightarrow{Route\ 25} 118$
	Yes	120.7	Twice	11.6	$100 \xrightarrow{Route\ 9} 11736 \xrightarrow{Walking} 13481 \xrightarrow{Walking} 10862 \xrightarrow{Route\ 25} 118$
13912	No	151.5	Twice	1.4	$13912 \xrightarrow{Route\ 17} 172 \xrightarrow{Walking} 12911 \xrightarrow{Route\ 6} 12707 \xrightarrow{Walking} 14295 \xrightarrow{Route\ 25} 118$
	Yes	203.6	Twice	1.8	$13912 \xrightarrow{Route\ 17} 12096 \xrightarrow{Walking} 13747 \xrightarrow{Route\ 19} 14206 \xrightarrow{Walking} 42 \xrightarrow{Route\ 25} 118$

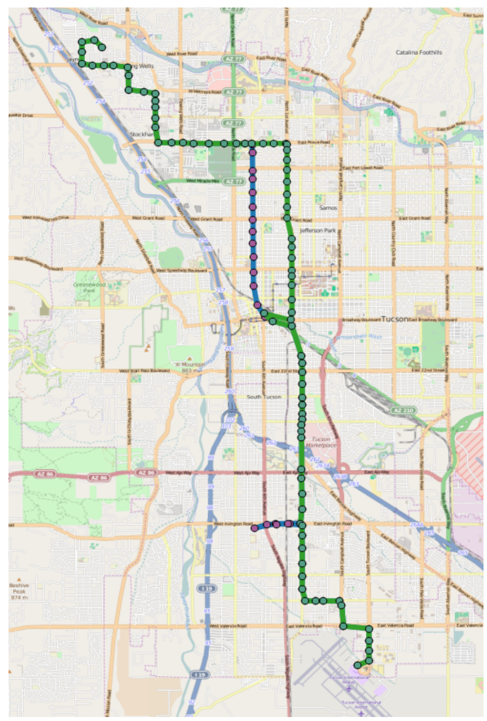
1) Both scenarios suggested that the optimal travel times became higher when the chance constraint was considered.

2) Walking was preferred in Scenario 1b because walking was more reliable than taking and waiting for buses. Both Routes 9 and 25 were chosen in Scenario 1. The major difference between Scenarios 1a and 1b was the mode selection to reach Stop 10862. Taking buses was chosen in Scenario 1a; while walking to the stop was chosen in Scenario 1b. The travel time of taking buses was usually shorter than walking time. However, congested traffic conditions may lead to less predictable and reliable bus arrival. The selection of walking may become an alternative to avoid

traffic congestion and ensure on-time arrival. Thus, walking became the optimal choice when considering chance constraints. The optimal path chosen for Stop 100 in scenario 1a and scenario 1b is shown in Figure 18 (left).

3) More reliable paths were chosen in Scenario 2. The differences of optimal travel times and walking time between Scenario 2a and Scenario 2b were minor (approximately 10 minutes and 1 minute, respectively). Route 6 chosen in Scenario 2a was planned on a busy roadway; whereas, Route 19 in Scenario 2b was planned on a roadway with relatively light traffic. Although the optimal travel time of Scenario 2a was slightly smaller than that of Scenario 2b, Scenario 2b could be a better path choice when considering a chance constraint with higher on-time arrival confidence level. The optimal path chosen for Stop 13912 in Scenario 2a and Scenario 2b is shown in Figure 18 (right).

Scenario 1



Scenario 2

FIGURE 18 Optimal paths with and without chance constraints

<u>Experiment II - Effects of Confidence Levels on Path Choice</u>

The second group of experiments was designed to investigate the effects of different on-time arrival confidence levels on path choice. The destination was again the Tucson International Airport. Three origins were selected, including the UA mall bus stop (Stop 100), Kain/Kimberly PI bus stop (Stop 13912), and 1st Av/Rillito Park (Stop 12900). Since traffic conditions greatly affect transit reliability and traffic congestion varied significantly between 6 am and 5 pm, transit service was considered reliable at 6 am and less reliable at 5 pm. Thus, six scenarios were created based on the three origins and these two TODs. Seven levels of on-time arrival confidence levels were tested for each scenario.

- Scenario 1a; origin: Stop 100; departure time: 6 am on weekday;
- Scenario 1b; origin: Stop 100; departure time: 5 pm on weekday;
- Scenario 2a; origin: Stop 13912; departure time: 6 am on weekday;
- Scenario 2b; origin: Stop 13912; departure time: 5 pm on weekday;
- Scenario 3a; origin: Stop 12900; departure time: 6 am on weekday;
- Scenario 3b; origin: Stop 12900; departure time: 5 pm on weekday;

Figure 19 and Table 3 show the optimal anticipated travel times for each scenario, and several findings are summarized below.

1) The optimal anticipated travel times increased with the increase in on-time arrival confidence level. For example, In Scenario 1, the optimal anticipated travel time was 82.45 minutes when a chance constraint was not considered. The optimal anticipated travel time increased to 109.3 minutes when the on-time arrival confidence level was set at 99.5%. The same trend can be observed in all of the scenarios. The trend was intuitive: for a fixed takeoff time, the more planning time, the higher the on-time arrival confidence level.

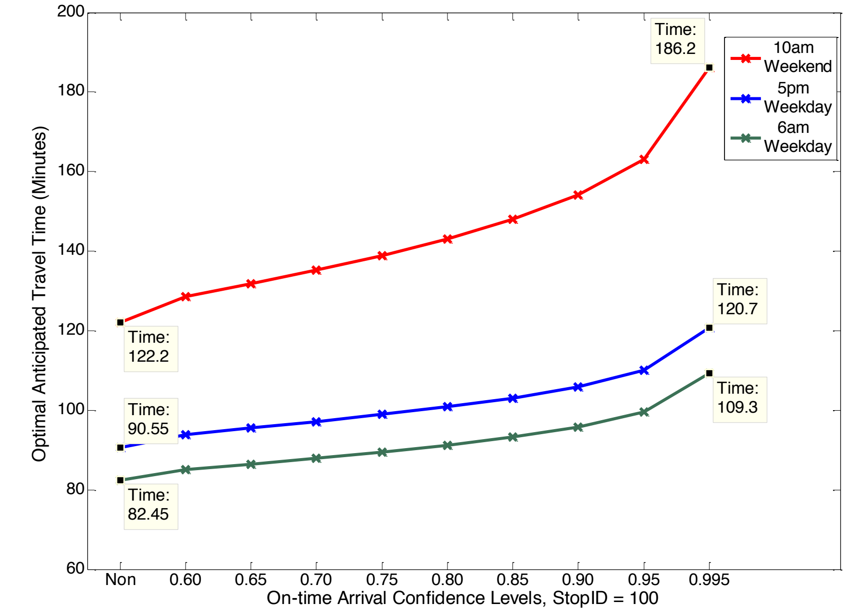
2) The optimal anticipated travel time at a given on-time arrival confidence level was greater at a departure time of 5 pm at 6 am. For example, without considering confidence level, the optimal anticipated travel times were 82.45 and 90.55 minutes, respectively. Generally, transit service was more reliable in the early morning than during peak hours.

3) Approximately 30% additional planning time could ensure on-time arrival at a relatively higher confidence level. Table 3 lists the optimal anticipated travel times when on-time arrival confidence level was not considered and at the 99.5% level for the six scenarios. Although the difference in time varied, the percentage differences suggested that trips could be on time at a reliability level 99.5% if 30% additional planning time was added as buffer time.

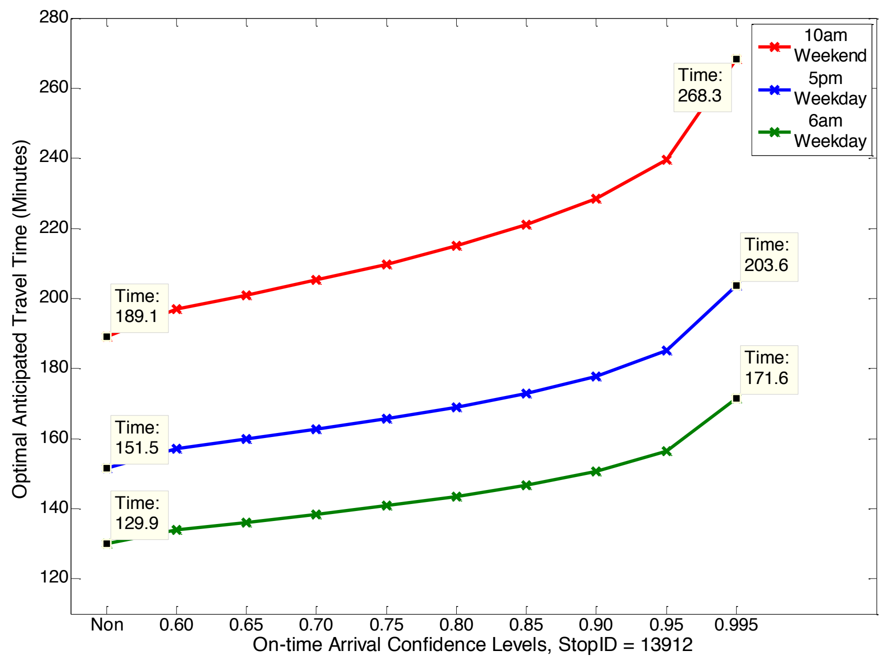
TABLE 3 Basic statistics

	Optimal anticipated travel time (minutes)		Difference	
	No Chance Constraint ($base$)	99.5% On-time Arrival Confidence Level	Minutes ($C_{0.995} - base$)	Percentage (%)

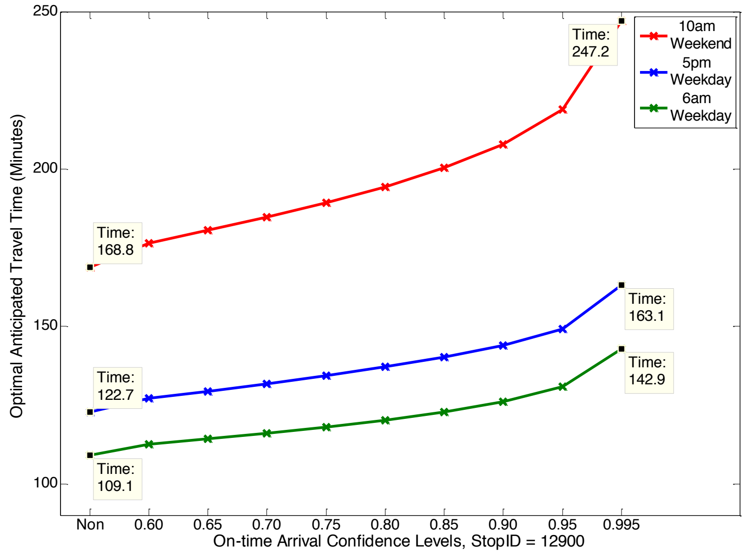
		$C_{0.995}$		$((C_{0.995} - base)/ base)$
Scenario 1a	82.45	109.3	26.85	32.57%
Scenario 1b	90.55	120.7	30.15	33.30%
Scenario 2a	129.9	171.6	41.7	32.10%
Scenario 2b	151.5	203.6	52.1	34.39%
Scenario 3a	109.1	142.9	33.8	30.98%
Scenario 3b	122.7	163.1	40.4	32.93%



Scenario 1



Scenario 2



Scenario 3

FIGURE 19 Optimal anticipated travel time vs. predefined confidence level

Experiment III - Weekend vs. Weekday

The third group of experiments was designed to investigate the optimal anticipated travel time and path choice on weekends and weekdays. Due to the light traffic on weekends, the transit service was presumed to be reliable and similar to weekday early morning service. However, Figure 19 shows that significant differences existed between the optimal anticipated travel time for 6 am on weekdays and 10 am on weekends. The major difference between the weekday timetable and the weekend timetable was the bus time headway. The time headway was typically set to be 10 or 15 minutes on weekdays; while it was as much as 60 minutes on weekends. Larger time headway resulted in longer waiting time at bus stops, and therefore the optimal anticipated travel time increased.

6. CONCLUSIONS

Transit systems are not thoroughly utilized in the U.S. Previous studies have shown that low fares and easily accessible transit information can convince increasing numbers of travelers to choose transit. With new technologies emerging, the ease of tracking and collecting transit fleet information in real-time helps improve both transit operations and real-time transit information quality. To further encourage travelers to take transit, an efficient decision tool would help traveler plan transit trips. In this study, a data-driven decision framework for intermodal trip planning was proposed and implemented. The advantages of the proposed framework are highlighted below:

- Both travel time and travel time reliability were considered in the system when planning intermodal travel. Travelers can be provided with two important transit measures, anticipated travel times and on-time arrival confidence levels to better plan their trips.
- The two transit measures were connected using a chance constrained decision model to obtain travel paths under different uncertainties. The chance constraint was transformed into an equivalent deterministic constraint based on the approximate normal distribution property of the path. Different random distributions could be included in the model and it's more easily implemented based on existing efficient shortest path algorithms.
- Walking mode was considered when transit passengers needed to transfer. Incorporating the walking mode into the system gave passengers more options regarding trip planning and helped passengers plan more reliable trips.

GTFS static and GTFS real-time data were collected and used for path optimization in Tucson, AZ and Boston, MA. Both data sets were utilized to estimate link travel time and travel time reliability in the constructed transit network. Three experiments under several different scenarios were conducted to study transit on-time arrival. The results of the three experiments suggested that:

- Optimal anticipated travel time increased with increasing on-time arrival confidence level. Essentially, more reliable planned transit paths usually involve longer anticipated travel times. As an example, approximately 30% additional time serves as a reference for allocating buffer time to ensure a high on-time arrival confidence level to the Tucson International Airport.
- It was found that walking was preferred instead of taking a transit detour. This is because the walking mode had relatively high reliability. The chance constrained decision model gave more weights to more reliable modes.

- Given different confidence level, different additional time are suggested as the buffer time to guarantee a higher on-time arrival confidence level during different traffic hours.

7. REFERENCES

[1] L. Redman, M. Friman, T. Gärling, and T. Hartig, "Quality attributes of public transport that attract car users: A research review," *Transp. Policy*, vol. 25, pp. 119–127, 2013.

[2] R. S. Ow and W. Jackson, "Use of Advanced Public Transportation Systems Is Described," *Newsline, Curr. Res. Public Transp. Transp. Res. Board*, vol. 24, no. 3.

[3] "UMTRI." [Online]. Available: http://www.umtri.umich.edu/. [Accessed: 07-Nov-2015].

[4] B. Ferris, K. E. Watkins, and A. Borning, "OneBusAway: Behavioral and Satisfaction Changes Resulting from Providing Real-Time Arrival Information for Public Transit," *Proc. CHI 2010*, pp. 1–15, 2010.

[5] J. Jariyasunant, D. B. Work, B. Kerkez, R. Sengupta, A. M. Bayen, and S. Glaser, "Mobile Transit Trip Planning with Real-Time Data," *Transp. Res. Board 89th Annu. Meet.*, no. September, pp. 1–17, 2010.

[6] "RideScout." [Online]. Available: http://www.ridescoutapp.com/. [Accessed: 07-Nov-2015].

[7] "TripPlanner." [Online]. Available: http://tripplanner.mta.info/MyTrip/ui_web/customplanner/TripPlanner.aspx. [Accessed: 07-Nov-2015].

[8] S. Tran, "Trip Planning." [Online]. Available: http://infoweb.suntran.com/hiwire?.a=iTripPlanning. [Accessed: 01-Aug-2015].

[9] "OneBusAway." [Online]. Available: http://onebusaway.org/. [Accessed: 01-Aug-2015].

[10] Y. Yan, Q. Meng, S. Wang, and X. Guo, "Robust optimization model of schedule design for a fixed bus route," *Transp. Res. Part C Emerg. Technol.*, vol. 25, pp. 113–121, 2012.

[11] Y. Shafahi and A. Khani, "A practical model for transfer optimization in a transit network: Model formulations and solutions," *Transp. Res. Part A Policy Pract.*, vol. 44, no. 6, pp. 377–389, 2010.

[12] T. C. Lam and K. a. Small, "The value of time and reliability: Measurement from a value pricing experiment," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 37, no. 2–3, pp. 231–251, 2001.

[13] M. L. Tam, W. H. K. Lam, and H. P. Lo, "Modeling Air Passenger Travel Behavior on Airport Ground Access Mode Choices," *Transportmetrica*, vol. 4, no. 2, pp. 135–153, 2008.

[14] C. Carrion and D. Levinson, "Value of travel time reliability: A review of current evidence," *Transp. Res. Part A Policy Pract.*, vol. 46, no. 4, pp. 720–741, 2012.

[15] "What is GTFS." [Online]. Available: https://developers.google.com/transit/gtfs/.

[16] "What is GTFS-realtime." [Online]. Available: https://developers.google.com/transit/gtfs-realtime/.

[17] "Protocol Buffers." [Online]. Available: https://developers.google.com/protocol-buffers/.

[18] S. Yang, E. Mechanics, A. Malik, and Y. Wu, "Travel Time Reliability using Hasofer Lind-Rackwitz Fiessler Algorithm and Kernel Density Estimation," pp. 12–15, 2014.

[19] J. Zhang, F. Liao, T. Arentze, and H. Timmermansa, "A multimodal transport network model for advanced traveler information systems," *Procedia Comput. Sci.*, vol. 5, pp. 912–919, 2011.

[20] O. J. Ahuja RK, Magnanti TL, *Network flows:theory, algorithms, and applications*. Prentice Hall, NJ, 1993.

[21] H. Noh, M. Hickman, and A. Khani, "Hyperpaths in Network Based on Transit Schedules," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2284, no. -1, pp. 29–39, 2012.

[22] A. Khani, M. Hickman, and H. Noh, "Trip-Based Path Algorithms Using the Transit Network Hierarchy," *Networks Spat. Econ.*, 2014.

[23] X. Su, H. Cai, M. Asce, B. T. Luong, and S. Ukkusuri, "From a Link-Node-Based Network Representation Model to a Lane-Based Network Representation Model : Two-Dimensional Arrangements Approach," *J. Comput. Civ. Eng.*, vol. 29, no. 3, pp. 1–11, 2015.

[24] X. Fu and W. H. K. Lam, "A network equilibrium approach for modelling activity-travel pattern scheduling problems in multi-modal transit networks with uncertainty," *Transportation (Amst).*, vol. 41, no. 1, pp. 37–55, 2014.

[25] "NetworkX Overview." [Online]. Available: https://networkx.github.io/#.

[26] "pymssql." [Online]. Available: http://www.pymssql.org/en/latest/.

[27] D. Olson, David L./Wu, *Enterprise Risk Management Models*. Springer Berlin Heidelberg, 2010.

[28] U. G. Acer, P. Giaccone, D. Hay, G. Neglia, and S. Tarapiah, "Timely data delivery in a realistic bus network," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1251–1265, 2012.

[29] S. Chandra and A. K. Bharti, "Speed Distribution Curves for Pedestrians During Walking and Crossing," *Procedia - Soc. Behav. Sci.*, vol. 104, pp. 660–667, 2013.

[30] D. M. L. Methode and D. Mlm, *Hoofdstuk 3 schatters en schattingen 1.* .

[31] F. Li, T. Wu, A. Badiru, M. Hu, and S. Soni, "A single-loop deterministic method for reliability-based design optimization," *Eng. Optim.*, vol. 45, no. 4, pp. 435–458, 2013.

[32] L. Liu, J. Yang, H. Mu, X. Li, and F. Wu, "Exact algorithms for multi-criteria multi-modal shortest path with transfer delaying and arriving time-window in urban transit network," *Appl. Math. Model.*, vol. 38, no. 9–10, pp. 2613–2629, May 2014.

[33] Z. Ji, Y. S. Kim, and A. Chen, "Multi-objective α-reliable path finding in stochastic networks with correlated link costs: A simulation-based multi-objective genetic algorithm approach (SMOGA)," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1515–1528, 2011.

[34] J. Hershberger, M. Maxel, and S. Sur, "Finding the k Shortest Simple Paths: A New Algorithm and its Implementation," *ACM Trans. Algorithms*, vol. 3, no. 4, pp. 45–56, 2007.

[35] W. Hoffman and R. Pavley, "A Method for the Solution of the Nth Best Path Problem," *J. ACM*, vol. 6, no. 4, pp. 506–514, 1959.

[36] J. Y. Yen, "Finding the K Shortest Loopless Paths in a Network," *Manage. Sci.*, vol. 17, no. 11, pp. 712–716, 1971.

[37] "wxPython." [Online]. Available: http://www.wxpython.org/. [Accessed: 01-Jan-2015].

[38] "Keyhole Markup Language." [Online]. Available: https://developers.google.com/kml/documentation/?hl=en. [Accessed: 01-Jan-2015].

[39] "simplekml." [Online]. Available: https://pypi.python.org/pypi/simplekml/. [Accessed: 01-Jan-2015].